10/531010 #2

PCT/IB 03/04326

01.10.03

Europäisches Patentamt

European Patent Office

Office européen des brevets

REC'D 10 OCT 2003

WIPO PCT

Bescheinigung

Certificate

Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. **Patent application No.** **Demande de brevet n°**

02292551.5

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

Der Präsident des Europäischen Patentamts; Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets p.o.

R C van Dijk

BEST AVAILABLE COPY

Anmeldung Nr:
Application no.: 02292551.5
Demande no:

Anmeldetag:
Date of filing: 16.10.02
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Koninklijke Philips Electronics N.V.
Groenewoudseweg 1
5621 BA Eindhoven
PAYS-BAS

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

Drift-free video encoding and decoding method

In Anspruch genommene Prioriät(en) / Priority(ies) claimed /Priorité(s) revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

H04N7/24

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR IE IT LI LU MC NL PT SE SK TR

"DRIFT-FREE VIDEO ENCODING AND DECODING METHOD"

FIELD OF THE INVENTION

The present invention relates to an encoding method for the compression of an original video sequence divided into successive groups of frames (GOFs), and to a corresponding video decoding method.

BACKGROUND OF THE INVENTION

Current video standards (from MPEG-1 to H.26 L) often use so-called hybrid solutions based on a predictive scheme where each frame is either intra coded (I frames) or temporally predicted from a given reference frame (the prediction options being, as shown in Fig.1, a forward prediction, for the P frames, or a bi-directional prediction, for the B frames), the prediction error thus obtained being then spatially transformed (a 2D-DCT transform is used in the standard schemes) to get advantage of spatial redundancies. According to a different approach proposed in the document "Three-dimensional subband coding of video", C.Podilchuk and al., IEEE Transactions on Image Processing, vol.4, n°2, February 1995, pp.125-139, a group of frames (GOF) is processed as a three-dimensional (2D+t, or 3D) structure and spatio-temporally filtered in order to compact the energy in the low frequencies (further studies included Motion Compensation in this scheme in order to improve the overall coding efficiency). The obtained 3D subband structure is depicted in Fig.2. The well known SPIHT algorithm, extended from 2D to 3D, was then used in order to efficiently encode the final coefficient bit-planes with respect to the spatio-temporal decomposition structure.

According to the usual implementation of a 3D subband codec, a motion-compensated (MC) spatio-temporal analysis is applied at the full original resolution, spatial scalability being then achieved by getting rid of the portions of the bitstream corresponding to the highest spatial subbands of the decomposition. When a motion compensation is used in the 3D analysis scheme, this method does not allow a perfect reconstruction of the video sequence at lower resolution, even with an infinite bit-rate. This phenomena will be referred to as drift in the following description. As explained in the document "Multiscale video compression using wavelet transform and motion compensation", P.Y.Cheng and al., Proceedings of the International Conference on Image Processing (ICIP95), Vol.1, 1995, pp.606-609, said drift comes from the order of wavelet transform and motion compensation that is not interchangeable. When a frame (A) is synthesized at a lower resolution (a), the following operation is applied :

$$\begin{aligned} a &= DWT_L(L) + MC[DWT_L(H)] \\ &= DWT_L(A) + [MC[DWT_L(H)] - DWT_L(MC[H])] \qquad (1) \end{aligned}$$

where DWT_L denotes the resolution downsample using the same wavelet filters as in the 3D analysis. In a perfect scalable solution, one wants to have:

$$a = DWT_L(A) \qquad (2)$$

The remaining part of the expression (1) therefore corresponds to the drift. It can be noticed that, if no MC is applied, the drift is removed. The same phenomena happens (except at the image borders) if a unique motion vector is applied to the frame. Yet, it is known that MC is unavoidable to achieve a good coding efficiency, and the likelihood of a unique global motion is small enough to eliminate this particular case in the following paragraphs.

Some authors, such as J.W.Woods and al in the document "A resolution and frame-rate scalable subband/wavelet video coder", IEEE Transactions on Circuits and Systems for Video Technology, vol.1, n°9, September 2001, pp.1035-1044, have already proposed technical solutions in order to get rid of this drift. However, in said document, the described scheme, in addition to being quite complex, implies the sending of an extra information (the drift correction necessary to correctly synthesize the upper resolution) in the bitstream, thus wasting some bits. The solution described in the document "Multiscale video compression..." previously cited avoids this bottleneck but works on a predictive scheme and is not transposable to the 3D subband codec.

It has then been proposed, in the European patent application n°02290155.7 (PHFR020002) filed on January 22nd, 2002, a solution avoiding these drawbacks. According to that solution, the video encoding method, used for the compression of an original video sequence divided into successive groups of frames (GOFs), comprised the steps of :

(1) generating from the original video sequence, by means of a wavelet decomposition, a low resolution sequence including successive low resolution GOFs ;

(2) performing on said low resolution sequence a low resolution decomposition, by means of a motion compensated spatio-temporal analysis of each low resolution GOF ;

(3) generating from said low resolution decomposition a full resolution sequence, by means of an anchoring of the high frequency spatial subbands resulting from the wavelet decomposition to said low resolution decomposition ;

(4) coding said full resolution sequence and the motion vectors generated during the motion compensated spatio-temporal analysis, for generating an output coded bitstream.

Said solution, in which the global structure of the decomposition tree in the 3DS analysis is preserved and no extra information is sent to correct the drift effect (only the decomposition/reconstruction mechanism is changed), is now recalled in a more detailed manner with reference to the coding scheme of Fig.3.

Two main steps are provided : (a) a motion compensation step at the lowest resolution, (b) an encoding step of the high spatial subbands. First, in order to avoid drift at lower resolutions, Motion Compensation (MC) was applied at this level. Consequently one first downsizes the GOF using the wavelet filters. Then the usual 3D subband MC-decomposition scheme is applied to this downsized GOF (it may be noticed that a side effect of this method is the reduction of the amount of motion vectors to be sent in the bitstream if the block size of the MC is the same as in a full-resolution process, which saves up some bits for texture coding). Before transmitting the subbands to a tree-based entropy coder (for instance to a 3D-SPIHT encoder such as described in the document "Low bit-rate scalable video coding with 3D set partitioning in hierarchical trees (3D-SPIHT)", B.J. Kim and al. IEEE Transactions on Circuits and Systems for Video Technology, vol.10, n°8, December 2000, pp.1374-1387), one puts the high spatial subbands that allow the reconstruction of the full resolution. The final tree structure looks very similar to that of a 3D subband codec such as the one described in the document "A fully scalable 3D subband video codec", V.Bottreau and al. Proceeding of IEEE Conference on Image Processing (ICIP2001), vol.2, pp.1017-1020, Thessaloniki, Greece, October 7-10, 2001, and so a tree-based entropy coder can be applied on it without any restriction.

Concerning the second step of coding the high spatial subbands, two main solutions are proposed, the first one without MC, and the second one with MC.

In the first solution, the high subbands simply correspond to the high frequency spatial subbands of the original (full resolution) frames of the GOF in the wavelet decomposition. Those subbands allow the reconstruction at full resolution at the decoding side. Indeed, the frames can be decoded at the low resolution. However, these frames correspond to the low spatial subband in the wavelet analysis of the original frames. Hence one has merely to put the low resolution frames and the corresponding high subbands together and apply a wavelet synthesis to obtain the full resolution frames, and thus to optimize the 3D-SPIHT encoder. In a MC scheme for a 3D subband encoder, the low temporal subbands always look like one of the original frames of the GOF. As a matter of fact :

$$L = \frac{1}{\sqrt{2}} \; [A + MC(B)] \qquad (3)$$

so L looks like A. Consequently, the high spatial subband of A should be placed with the low resolution decomposition corresponding to L. This approach (reordering of the high spatial subbands in the case of forward MC) is illustrated in Fig.4, where jt indicates the temporal decomposition level (0 for the full-frame rate, jt_max for the lowest frame rate), nf is the subband index at the temporal level jt, DWT_H denotes the high frequency wavelet filter and the coefficients c_{jt} are multiplication coefficients.

In the second solution, as using MC in every subband does not allow a reconstruction with no drift, it is also possible to partially use MC to construct the high spatial subbands and still be able to reconstruct every resolution. Instead of directly using the high frequency spatial subbands of the wavelet decomposition, a wavelet decomposition is carried out on a prediction error obtained from the MC performed on the full resolution sequence and reusing for instance the motion vectors of the low resolution.

SUMMARY OF THE INVENTION

It is then an object of the invention to improve the previously described solution by keeping its good behaviour at low resolution while getting closer to the performance of a classic 3D subband codec at full resolution.

To this end, the invention relates to a video encoding method for the compression of an original video sequence divided into successive groups of frames (GOFs), said method comprising the steps of :

(1) generating from the original video sequence, by means of a wavelet decomposition, a set of low resolution frames organized in successive low resolution GOFs ;

(2) performing on said low resolution frames a motion compensated spatio-temporal analysis, leading to a low resolution sequence ;

(3) performing a motion compensated spatio-temporal analysis of each full resolution GOF of the original video sequence ;

(4) replacing at each temporal decomposition level the low-frequency subbands of said decomposition by the corresponding spatio-temporal subbands of the low resolution sequence ;

(5) coding the modified sequence thus obtained and the motion vectors generated during the motion compensated spatio-temporal analysis of each full resolution GOF, for generating an output coded bitstream.

The invention also relates to a corresponding decoding method.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will now be described in a more detailed manner, with reference to the accompanying drawings in which :

- Fig.1 illustrates the different predictions in a typical hybrid video encoding scheme ;
- Fig.2 shows a 3D subband decomposition ;
- Fig.3 depicts an embodiment of an encoding scheme according to a previous embodiment ;
- Fig.4 illustrates the reordering of the high spatial subbands (for a forward motion compensation) ;

- Fig.5 illustrates the main steps of the encoding method according to the invention ;

- Fig.6 illustrates the corresponding motion compensated temporal filtering decomposition scheme ;

- Fig.7 illustrates at the decoding side an implementation of a synthesis scheme corresponding to the encoding method of Fig.5.

DETAILED DESCRIPTION OF THE INVENTION

As for the previously described solution, the present invention is now explained with reference to its basic steps : (a) motion compensation at the lowest resolution (this first step, Motion Compensation (MC), is, in fact, strictly equivalent to the one described in the case of the previous solution : one first downsizes the GOF using the spatial wavelet filters, and the usual 3D subband MC-decomposition scheme is then applied to this downsized GOF), (b) encoding the high spatial subbands.

The main difference with said previous solution lies in the second step, the principle of which is to inject at each decomposition level the temporal subbands of the low spatial resolution analysis into those of the full-resolution one. It is thus possible to reconstruct the original frames at the decoder side while performing a real temporal filtering (and not just an intra coding or a predictive difference – as in the previous solution - for the high frequency spatial subbands).

The following equations explain the mechanism in a more detailed manner. As said above, the first temporal analysis is performed at low resolution, which may be expressed by the equations (4) and (5) :

$$H_d = [B_d - MC_{down}(A_d)]/\sqrt{2} \qquad (4)$$

$$L_d = [\sqrt{2} * A_d + MC_{down}^{-1}(H_d)] \qquad (5)$$

with the following notations:

A = reference frame

B = current frame

DWT = discrete wavelet transform

A_d = low-frequency spatial subband of the DWT of frame A, i.e. a low-spatial resolution version of frame A

B_d = low-frequency spatial subband of the DWT of frame B, i.e a low-spatial resolution version of frame B

H = high-frequency temporal subband at the low spatial resolution

L = low-frequency temporal subband at the low spatial resolution

MC_{down} = motion compensation performed on low-resolution (i.e. sub-sampled) frames.

MC^{-1} = inverse motion compensation (motion vectors computed to predict a frame B from a frame A are reversely used to predict the frame A from the frame B).

The equations (6) to (9) then allow to define L_s and H_s :

$$H' = B - MC_{full}(A) \quad (6)$$

$$L' = \sqrt{2} * A + MC_{full}^{-1}(H) \quad (7)$$

$$H_s = H' \quad (8)$$

$$L_s = \sqrt{2} . L' \quad (9)$$

with :

X_s = union of the three high-frequency spatial subbands of the DWT of a given frame X (with $X_S = H_S$ or L_S)

MC_{full}= motion compensation performed on full-resolution frames

L' and H' = respectively the low-frequency and high-frequency temporal subbands in a conventional 3D subband scheme

$H = DWT^{-1} [H_d \cup H_s]$

$L = DWT^{-1} [L_d \cup L_s]$

Once all the low-frequency and high-frequency temporal subbands have been generated at a given temporal level jt, both at low and full spatial resolutions, the low-frequency temporal subbands L are further decomposed to achieve the next temporal level jt+1.

This is repeated at each step of the temporal decomposition, leading finally to a structure of the temporal decomposition which is very similar to that of a classic 3D subband encoder. The low frequency temporal subband of the last level and the high frequency temporal subbands of all levels are then spatially decomposed through wavelet filters and encoded to form the bitstream.

The described invention keeps the good behaviour of the previous solution at low resolution while getting closer to the performance of a classic 3D subband codec at full resolution (the global structure of the decomposition tree in the 3D subband analysis is preserved and no extra information is sent to correct the drift effect ; only the decomposition/reconstruction mechanism is changed). The main upgrade comes from the new approach to generate the high-frequency spatial subbands, that brings more coherence to the decomposition tree and therefore improves the coding efficiency of the system.

At the decoder, all the previous equations can be reverted to allow a good reconstruction. Only a ^ is added to every subband in order to indicate that decoding is now concerned and that some information might have been lost. First a classic 3D subband synthesis at low resolution allows to give back the low spatial resolution subbands A_d and B_d from L_d and H_d :

$$\hat{A}_d = \frac{1}{\sqrt{2}}\left[\hat{L}_d - MC_{down}^{-1}(\hat{H}_d)\right] \quad (10)$$

$$\hat{B}_d = MC_{down}(\hat{A}_d) + \sqrt{2} \cdot \hat{H}_d \quad (11)$$

It is also easy to get A_s by synthesizing H and by reverting the equation (7). The process is explained by the equations (12) to (15) :

$$\hat{H} = DWT^{-1}[\hat{H}_d \cup \hat{H}_s] \quad (12)$$

$$\hat{L} = DWT^{-1}[\hat{L}_d \cup \hat{L}_s] \quad (13)$$

$$\hat{A}''_s = \frac{1}{\sqrt{2}}\left[\hat{L} - MC_{full}^{-1}(\hat{H})\right] \quad (14)$$

$$\hat{A}_s = A''_s \quad (15)$$

Then $\hat{A}$ is simply reconstructed from $\hat{A}_d$ and $\hat{A}_s$. Consequently one can get B_s and finally synthesize B. This is summarized by the system of equations (16) to (19) :

$$\hat{A} = DWT^{-1}[\hat{A}_d \cup \hat{A}_s] \quad (16)$$

$$\hat{B}''_s = MC_{full}(\hat{A}) + \hat{H} \quad (17)$$

$$\hat{B}_s = \hat{B}''_s \quad (18)$$

$$\hat{B} = DWT^{-1}[\hat{B}_d \cup \hat{B}_s] \quad (19)$$

These operations are repeated until the very first temporal level, i.e. until the GOF is fully decoded. It can clearly be seen that this scheme generates no drift since perfect reconstruction is achieved as soon as L and H are completely transmitted in the bit-stream (it can also be noted that the full spatial resolution synthesis is now intimately linked with the low resolution one at each temporal level, which was not the case in the previous solution).

The encoding principle defined above will now be described in a more detailed manner, with reference to Fig.5, that illustrates the main steps of the encoding method, and Fig.6, that illustrates the corresponding motion compensated temporal filtering scheme.

In the encoding scheme of Fig.5, the original group of frames (this current GOF comprises full resolution frames) is first used for generating, by means of a wavelet decomposition, low resolution frames on which a motion compensated spatio-temporal analysis is then performed. A low resolution sequence is thus obtained. The original full resolution frames (i.e. each full resolution GOF) is also used for performing

a motion compensated spatio-temporal analysis (the corresponding successive steps are designated by : "MC-temporal analysis" and "wavelet decomposition").

After these two parallel sets of steps performed on the full resolution frames, the low frequency subbands of the decomposition thus obtained are iteratively replaced, at each temporal decomposition level, by the corresponding spatio-temporal subbands of the low resolution sequence, according to the following operations :

(a) first, a storing operation, for storing the high frequency spatio-temporal subbands of the decomposition in view of the final encoding step ;

(b) then a wavelet synthesis, performed from the low frequency spatio-temporal subbands of said decomposition ;

(c) then a test concerning the rank of the temporal decomposition level, for storing the low frequency spatio-temporal subbands of the decomposition if said level is the last one, the two parallel sets of steps being on the contrary further carried out for the next temporal level if said level is not the last one.

More detailed representations of the whole decomposition scheme and the corresponding motion-compensated synthesis scheme at the decoding side can be seen in Fig.6 and Fig.7 respectively. This example of a spatio-temporal decomposition according to the invention is related to a GOF of only four frames A0 to A3 (for the sake of simplicity), with a forward motion compensation and two decomposition levels. The high and low frequency (H'_0, H'_1 and L'_0, L'_1 respectively) temporal subbands are computed from the original frames by using the so-called lifting scheme, described for instance in the document "Factoring wavelet transforms into lifting steps", I.Daubechies and W.Sweldens, Bell Laboratories technical report, Lucent Technologies, 1996. The notations DWT and DWT^{-1} respectively designate the wavelet decomposition and the wavelet synthesis. The right side of Fig.6 illustrates successively the first spatio-temporal decomposition level, the inverse synthesis applied to the low frequency spatio-temporal subbands of the decomposition and the second spatio-temporal decomposition level (performed after the replacement of the low frequency subbands of the decomposition by the corresponding spatio-temporal subbands of the low resolution sequence, said replacement being indicated by the arrows coming from the left side of Fig.6).

CLAIMS :

1. A video encoding method for the compression of an original video sequence divided into successive groups of frames (GOFs), said method comprising the steps of :

(1) generating from the original video sequence, by means of a wavelet decomposition, a set of low resolution frames organized in successive low resolution GOFs ;

(2) performing on said low resolution frames a motion compensated spatio-temporal analysis, leading to a low resolution sequence ;

(3) performing a motion compensated spatio-temporal analysis of each full resolution GOF of the original video sequence ;

(4) replacing at each temporal decomposition level the low-frequency subbands of said decomposition by the corresponding spatio-temporal subbands of the low resolution sequence ;

(5) coding the modified sequence thus obtained and the motion vectors generated during the motion compensated spatio-temporal analysis of each full resolution GOF, for generating an output coded bitstream.

2. A video decoding method, provided for decoding a coded bitstream corresponding to a video sequence coded by means of a video encoding method comprising, for the compression of said original video sequence, the steps of :

(1) generating from the original video sequence, by means of a wavelet decomposition, a set of low resolution frames organized in successive low resolution GOFs ;

(2) performing on said low resolution frames a motion compensated spatio-temporal analysis, leading to a low resolution sequence ;

(3) performing a motion compensated spatio-temporal analysis of each full resolution GOF of the original video sequence ;

(4) replacing at each temporal decomposition level the low-frequency subbands of said decomposition by the corresponding spatio-temporal subbands of the low resolution sequence ;

(5) coding the modified sequence thus obtained and the motion vectors generated during the motion compensated spatio-temporal analysis, of each full resolution GOF, for generating an output coded bitstream ;

said video decoding method comprising the steps illustrated in the MC temporal synthesis scheme of Fig.7.

Abstract

Three-dimensional (3D) subband coding schemes use motion compensation in their temporal filtering stage. Unfortunately, this procedure introduces two drawbacks : (a) the MC being applied at the full resolution, a drift appears when decoding at a lower resolution, and (b) all the motion vectors estimated at full resolution are transmitted, which is a waste of bits. According to the invention, a low resolution sequence is first obtained by generating from the original input sequence of frames –by means of a wavelet decomposition- low resolution frames and performing on them a motion compensated spatio-temporal analysis. Then, a motion compensated spatio-temporal analysis of each full resolution group of frames is performed, and the low frequency subbands of the decomposition are finally replaced, at each temporal decomposition level, by the corresponding spatio-temporal subbands of the generated low resolution sequence. The modified sequence thus obtained is finally coded. Thanks to this approach, a good behaviour at low resolution is maintained (no more drift) while getting closer to the performance of a classic 3D subband codec at full resolution.

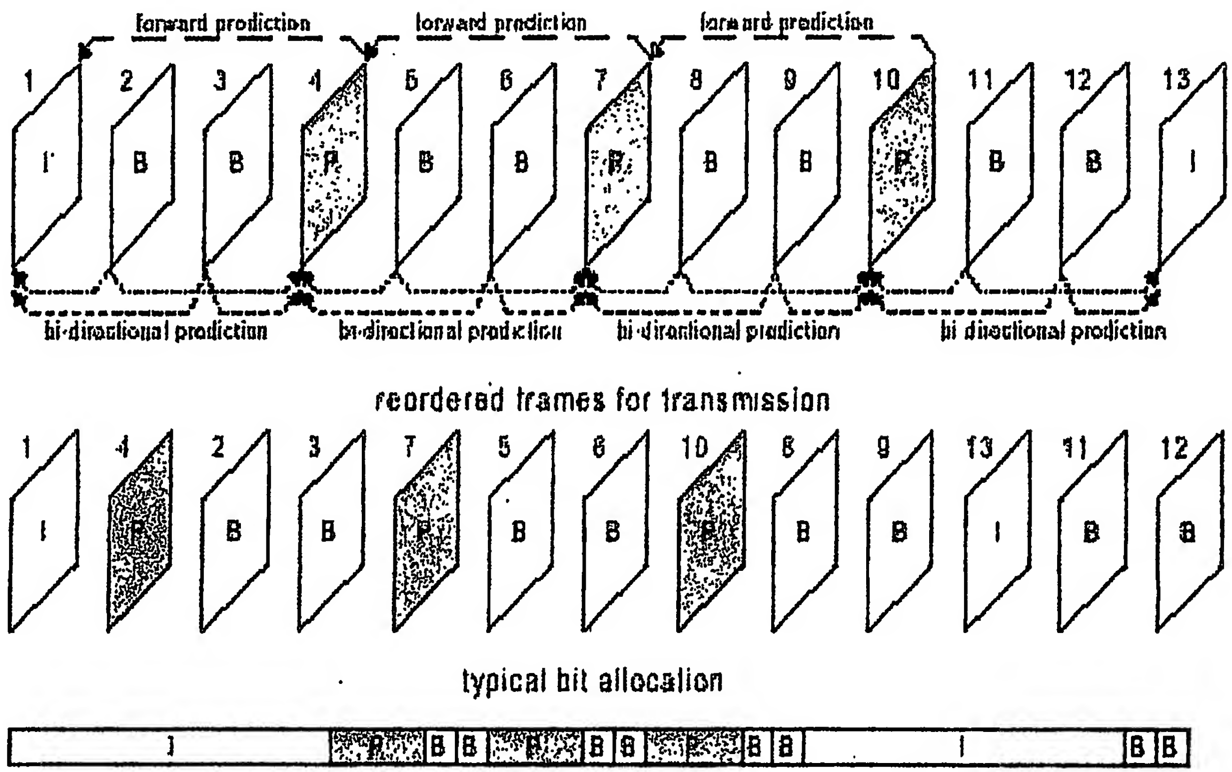

forward prediction
forward prediction
forward prediction
1
2
3
4
5
6
7
8
9
10
11
12
13
I
B
B
P
B
B
P
B
B
P
B
B
I
bi-directional prediction
bi-directional prediction
bi-directional prediction
bi-directional prediction
reordered frames for transmission
1
4
2
3
7
5
6
10
8
9
13
11
12
I
P
B
B
P
B
B
P
B
B
I
B
B
typical bit allocation
I
P
B
B
P
B
B
P
B
B
I
B
B

FIG.1

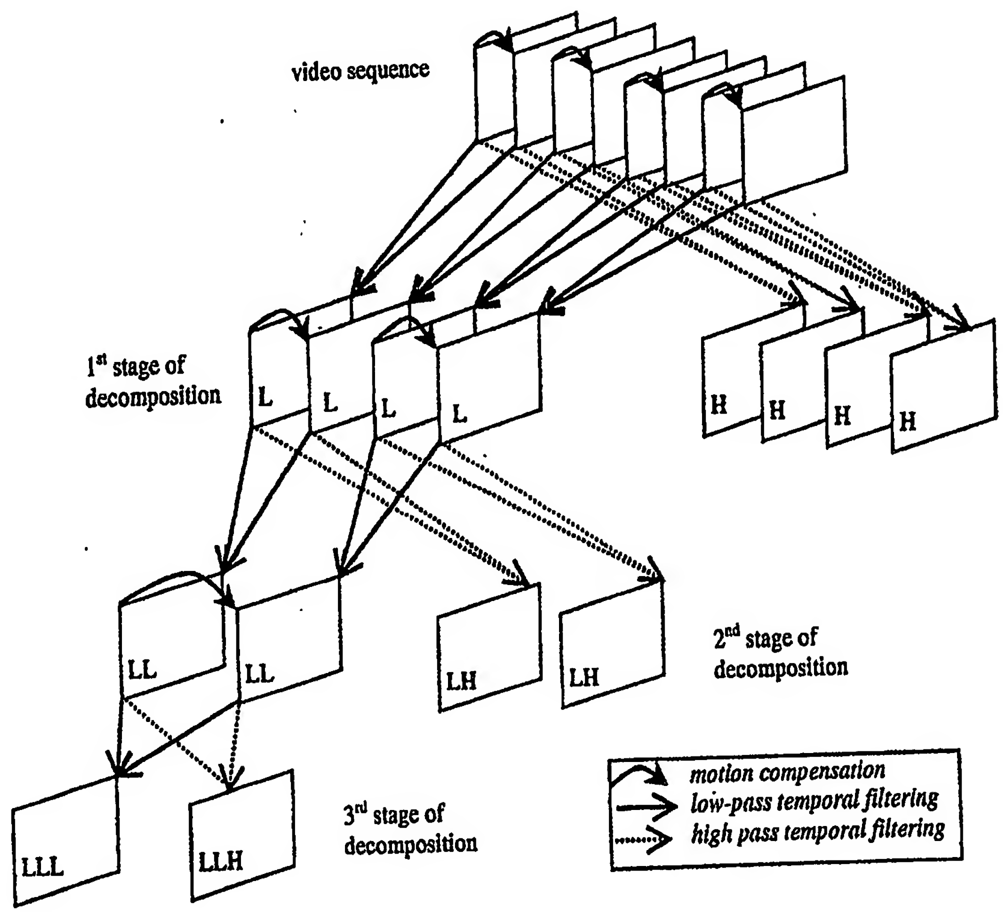

video sequence
1st stage of decomposition
L
L
L
L
H
H
H
H
LL
LL
LH
LH
2nd stage of decomposition
LLL
LLH
3rd stage of decomposition
motion compensation
low-pass temporal filtering
high pass temporal filtering

FIG. 2

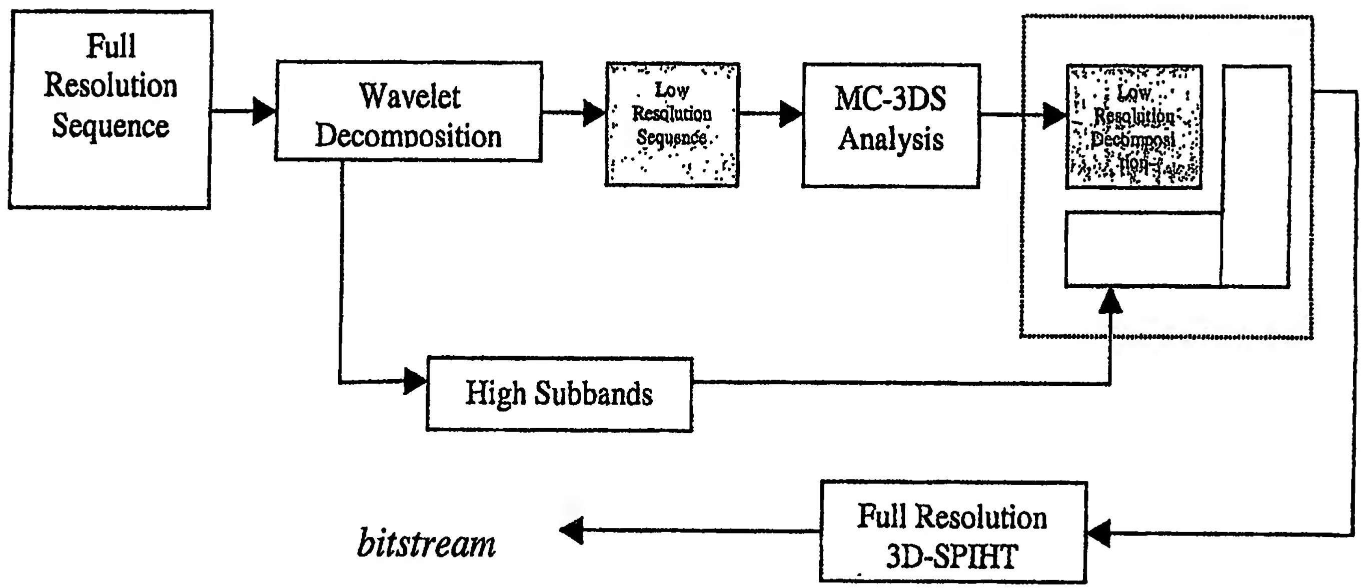
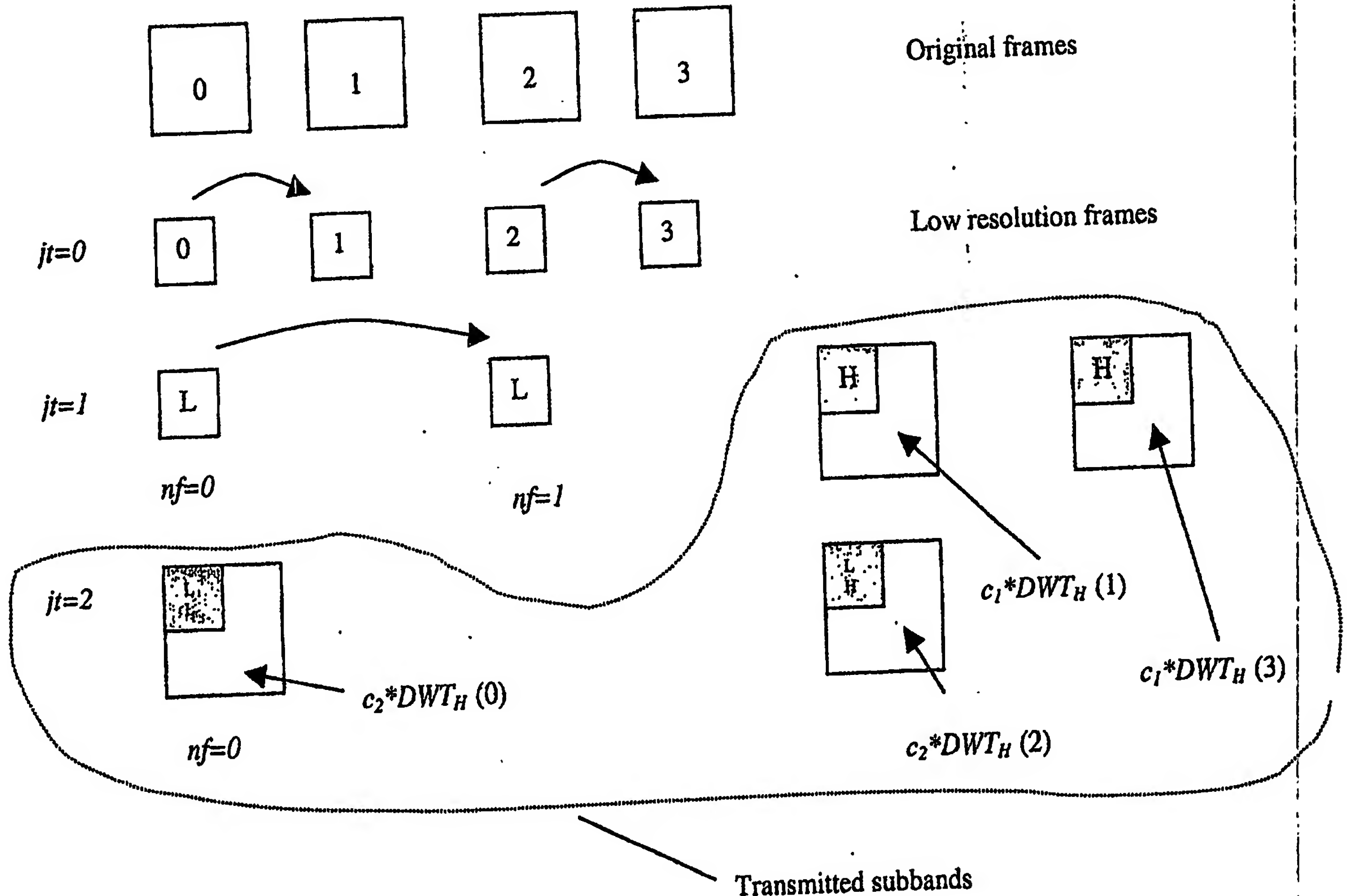

Full Resolution Sequence
Wavelet Decomposition
Low Resolution Sequence
MC-3DS Analysis
Low Resolution
High Subbands
Full Resolution 3D-SPIHT
bitstream

FIG.3



Original frames
Low resolution frames
0
1
2
3
jt=0
jt=1
jt=2
L
nf=0
nf=1
H
c_1*DWT_H (1)
c_1*DWT_H (3)
c_2*DWT_H (0)
c_2*DWT_H (2)
Transmitted subbands

FIG. 4

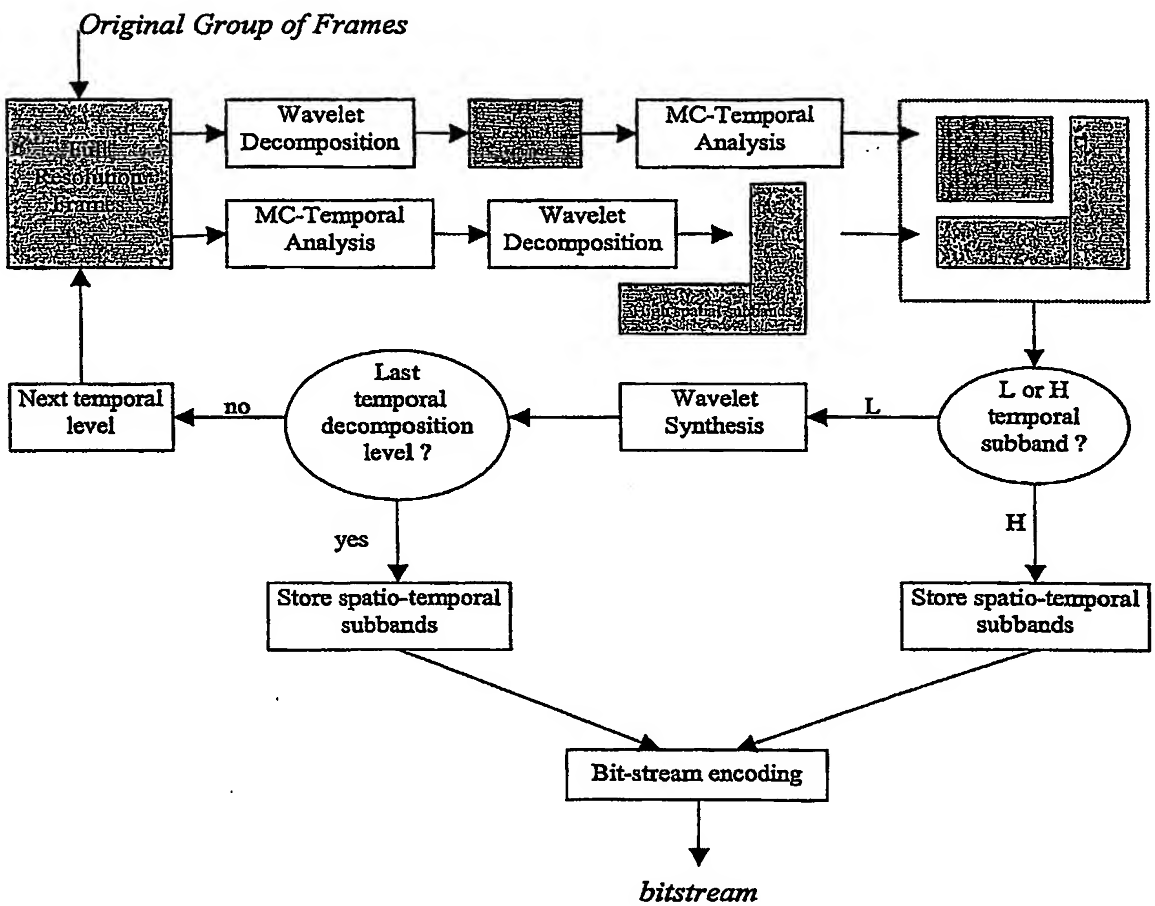
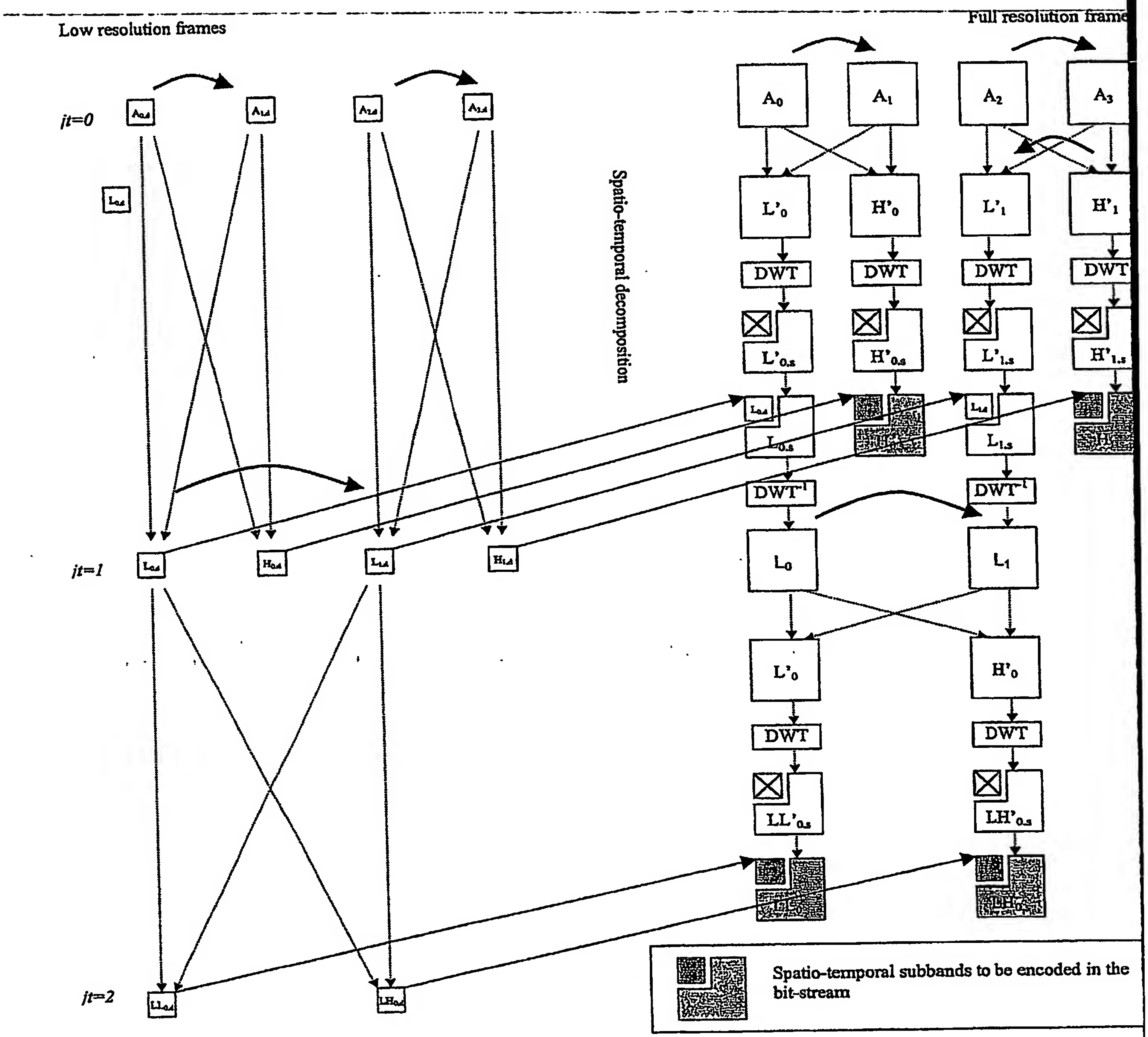

Original Group of Frames
Wavelet Decomposition
MC-Temporal Analysis
MC-Temporal Analysis
Wavelet Decomposition
Next temporal level
no
Last temporal decomposition level ?
Wavelet Synthesis
L
L or H temporal subband ?
yes
H
Store spatio-temporal subbands
Store spatio-temporal subbands
Bit-stream encoding
bitstream

FIG.5



Low resolution frames
Full resolution frame
jt=0
jt=1
jt=2
Spatio-temporal decomposition
A₀
A₁
A₂
A₃
L'₀
H'₀
L'₁
H'₁
DWT
L'0,s
H'0,s
L'1,s
H'1,s
L0,s
L1,s
DWT⁻¹
L₀
L₁
LL'0,s
LH'0,s
Spatio-temporal subbands to be encoded in the bit-stream

FIG. 6

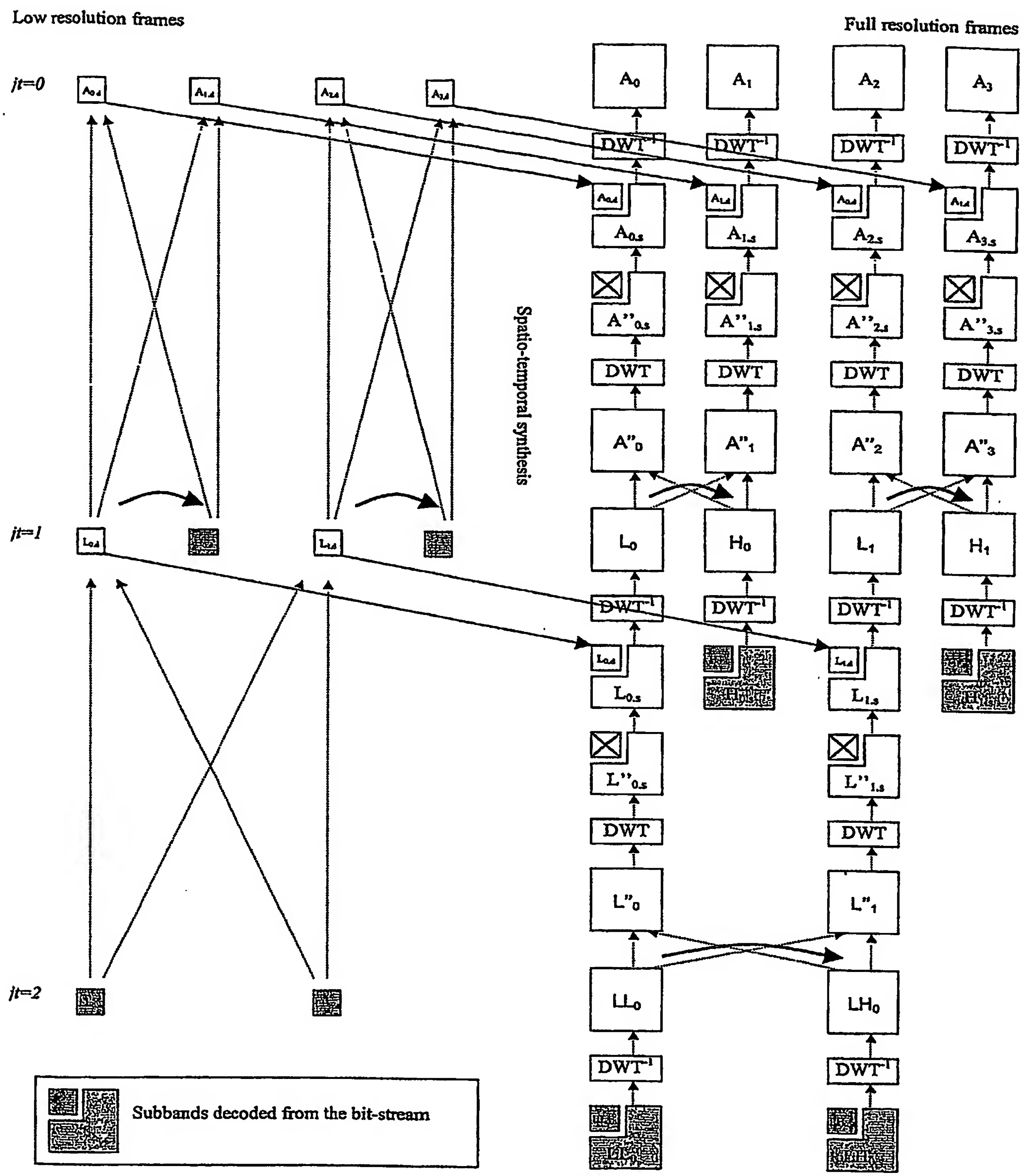

Low resolution frames
Full resolution frames
jt=0
jt=1
jt=2
Spatio-temporal synthesis
A0
A1
A2
A3
DWT-1
A0,s
A1,s
A2,s
A3,s
A''0,s
A''1,s
A''2,s
A''3,s
DWT
A''0
A''1
A''2
A''3
L0
H0
L1
H1
L0,s
L1,s
L''0,s
L''1,s
L''0
L''1
LL0
LH0
Subbands decoded from the bit-stream

FIG. 7